Multilingual Visual Sentiment Concept Matching







Nikolaos Pappas, Miriam Redi, Mercan Topkara, Brendan Jou, Hongyi Liu, Tao Chen, Shih-Fu Chang

IDIAP Yahoo JWPlayer Columbia University

ICMR 2016

Motivation

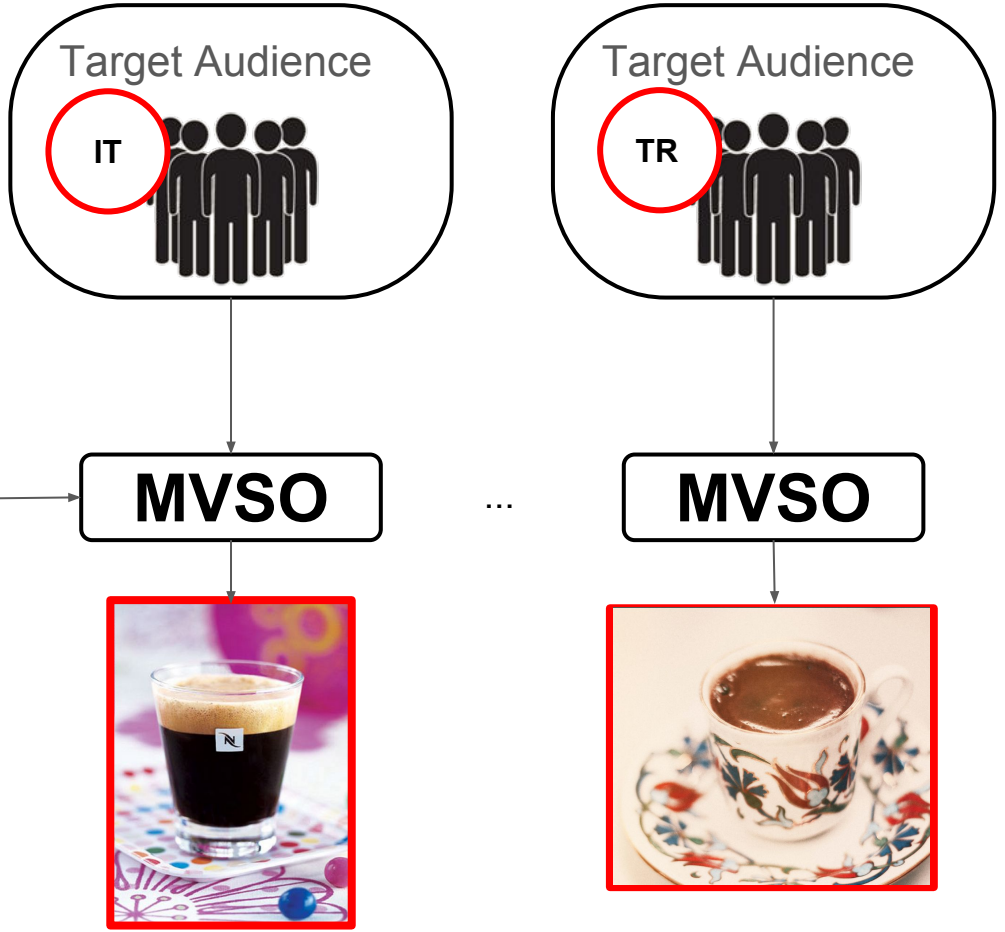
- How to analyze and retrieve multimedia data generated by a diverse, multicultural population?
- What are the lexical and visual differences of similar concepts across languages? How do different cultures use images to express sentiment and emotions?

Applications

Multilingual sentiment analysis of images

Applications

Target image selection based on cultural characteristics of the audience



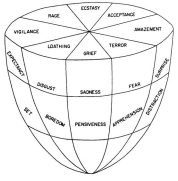
Challenges

- How to **collect** multilingual sentiment-biased images and metadata? **MVSO!**
- How do different languages **describe** visual emotions? **MVSO!**

- How to **compare and analyze visual concepts across languages**? **THIS WORK**

Brendan Jou, Tao Chen, Nikolaos Pappas, Miriam Redi, Mercan Topkara, Shih-Fu Chang
Visual Affect Around the World: A Large-scale Multilingual Visual Sentiment Ontology
ACM Multimedia 2015, Brisbane, Australia

Multilingual Visual Sentiment Ontology (MVSO)



EMOTION KEYWORDS
[Plutchik 1980]

FLICKR CRAWLING

ADJECTIVE NOUN PAIRS DISCOVERY

FREQUENT ANPs (automatic corpus)

FILTERING

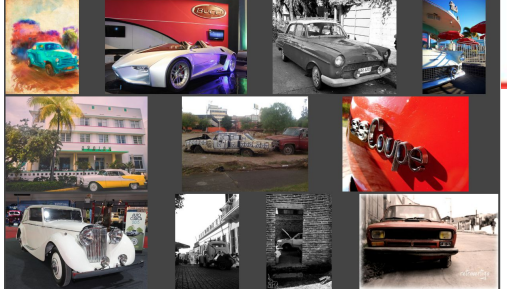
ORGANIZED COLLECTION OF MULTILINGUAL AFFECTIVE VISUAL CONCEPTS: ANPs

ADJECTIVE-NOUN PAIRS

Affective content, 12 languages, semantically consistent

ANP = ADJECTIVE NOUN PAIR

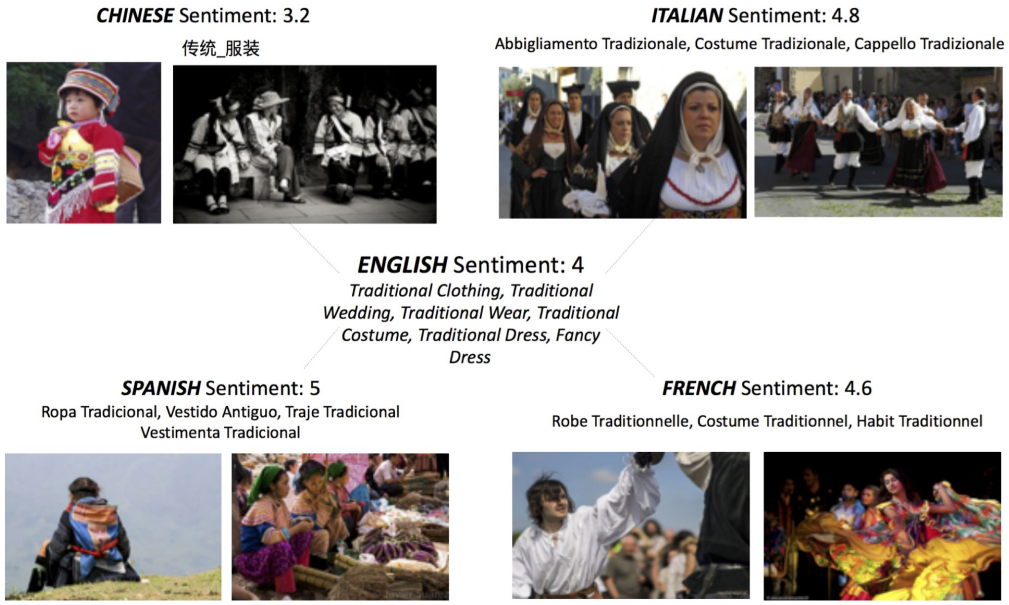
old cars, classic cars,..



Brendan Jou, Tao Chen, Nikolaos Pappas, Miriam Redi, Mercan Topkara, Shih-Fu Chang
Visual Affect Around the World: A Large-scale Multilingual Visual Sentiment Ontology
ACM Multimedia 2015, Brisbane, Australia

Discovering Multilingual Clusters

- Cultural insights based on semantically related concepts
- Each cluster reveals
  - Wording variation
  - Sentiment variation
  - Visual content variation

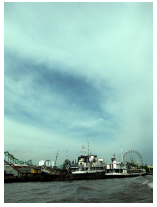
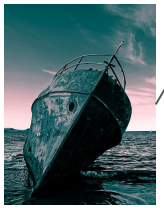


Example: Western vs. Eastern languages

FRENCH: bateaux abandones (abandoned boats **sent:1.2**)


ENGLISH: old boats **sent:1.7**

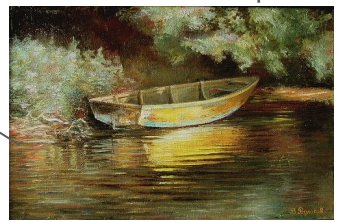

SPANISH: barco abandonado (abandoned boat **sent:1.0**)

CLUSTER:
OLD BOAT
ABANDONED BOAT
ABANDONED SHIP

CHINESE: 旧 船 (old boats, **sent:2.8**)


RUSSIAN: старая лодка (old boat, **sent:1.7**)




Example: Culturally-unique clusters

- Cultural insights based on distinctive concepts
- Each cluster reveals
  - Uniqueness
  - Expressivity
  - Cultural specificity

Proposed Framework

1. Translate each original ANP into English
2. Use word embeddings to convert ANPs to vectors and cluster

MVSO Concepts

ENGLISH
old boats
happy dog
healthy coffee
...

FRENCH
bateaux abandones (abandoned boats)
chien heureux (happy dog)
corps sain (healthy body)
...

GERMAN
verlassene gebäude (abandoned building)
glücklicher hund (happy dog)
gesunde ernährung (healthy eating)
...

ITALIAN
casa abbandonata (abandoned building)
cane divertente (funny dog)
cibo_sano (healthy food)
...

DUTCH
oude gebouw (old building)
mooie kat (lovely cat)
...

SPANISH
barco abandonado (abandoned boat)
perro feliz (happy dog)
desayuno saludable (healthy breakfast)
...

CHINESE
旧 船 (old boat)
可爱 狗 (cute dog)
健康 生活 (healthy lifestyle)
...

RUSSIAN
старая лодка (abandoned building)
здоровое питание (healthy eating)
...

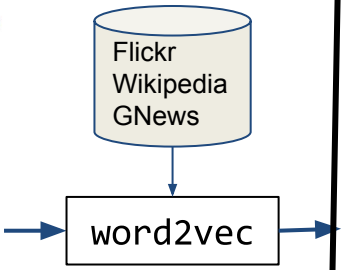
TURKISH
eski evler (old houses)
...

POLISH
stary budynek (old building)
...

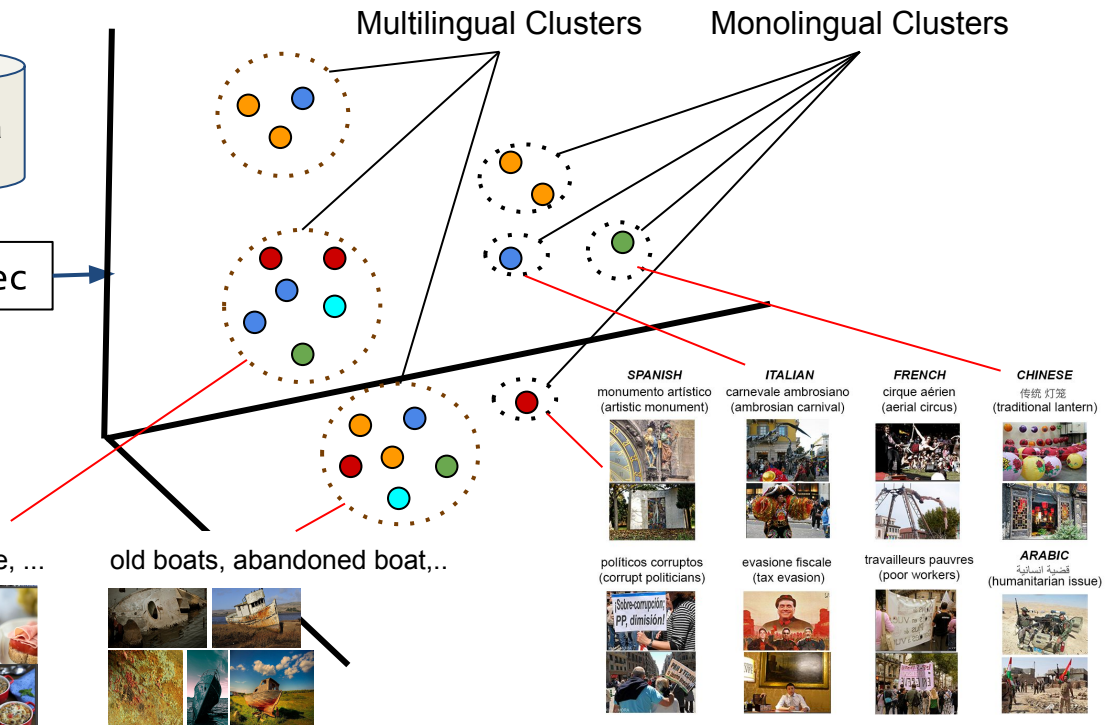
ARABIC
قصر قديم (abandoned palace)
...

PERSIAN
بیت مهجور (abandoned house)
...

Concept Matching



Concept Clustering



DATA

Multilingual Visual Sentiment Ontology (MVSO) Data

- 7.36M+ Flickr images
- ~16K affective visual concepts: Adjective-Noun Pairs (ANPs)
- Co-occurrence (emotion, ANP)
- Sentiment value (text-based)
- 12 languages detected

Italian

Treno storico

Bella giornata

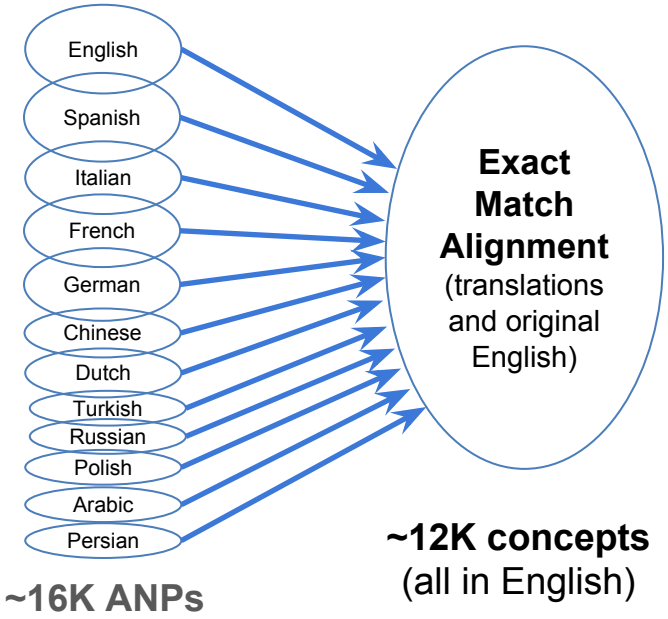
Treno veloce

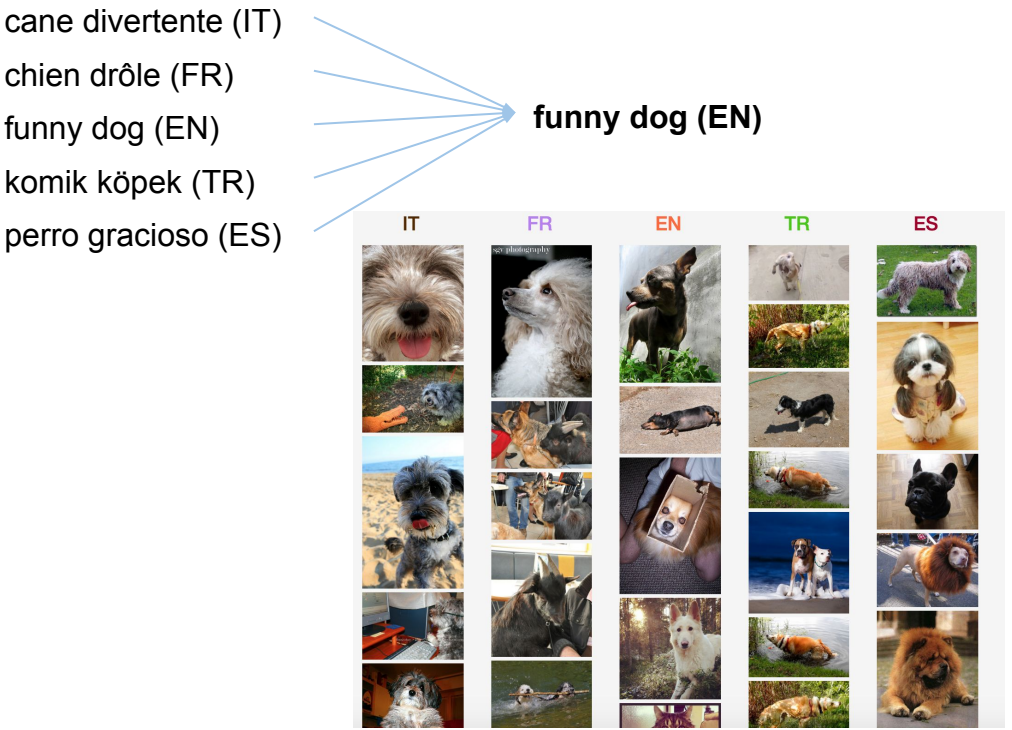
Language	Concepts	Images
English	4421	447997
Spanish	3381	37528
Italian	3349	25664
French	2349	16807
Chinese	504	5562
German	804	7335
Dutch	348	2226
Russian	129	800
Turkish	231	638
Polish	63	477
Persian	15	34
Arabic	29	23

CONCEPT MATCHING

Exact Concept Matching with English Translation

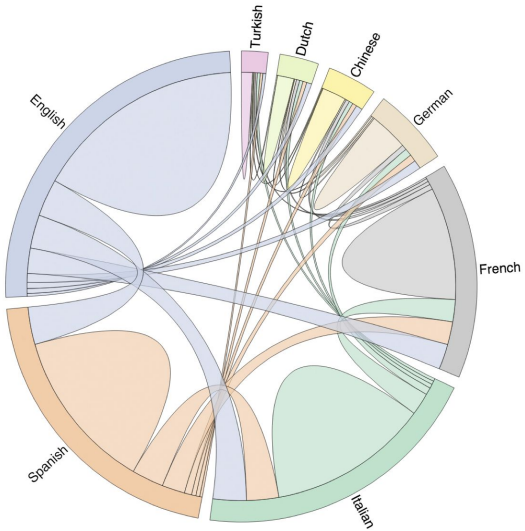
Reflection of what we would see depending solely on translation to understand other cultures and their interpretation of concepts *(wedding, new year, traditional costumes)*



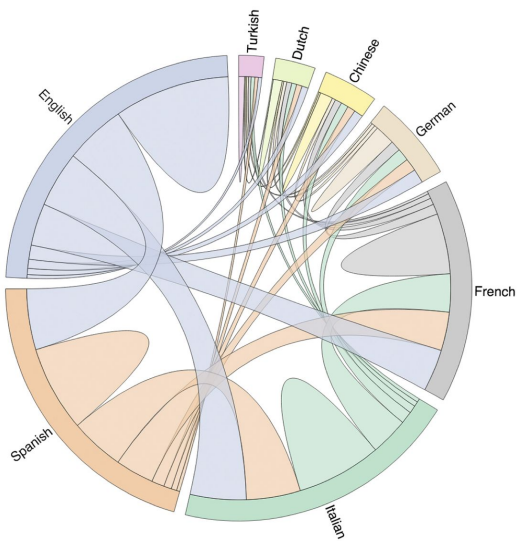


Limitations of Exact Concept Matching

- Low ratio of crosslingual related concepts
  - 9.8K ANPs in monolingual clusters with exact matching based alignment
  - Number of monolingual clusters was below 2.5K with all approximate matching clustering methods

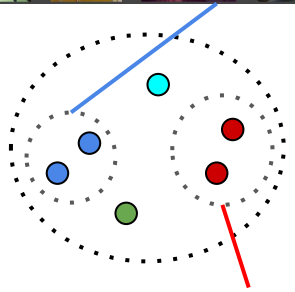


(a) Exact matching



(b) Approximate matching

SPANISH: desayuno saludable (**healthy breakfast**)

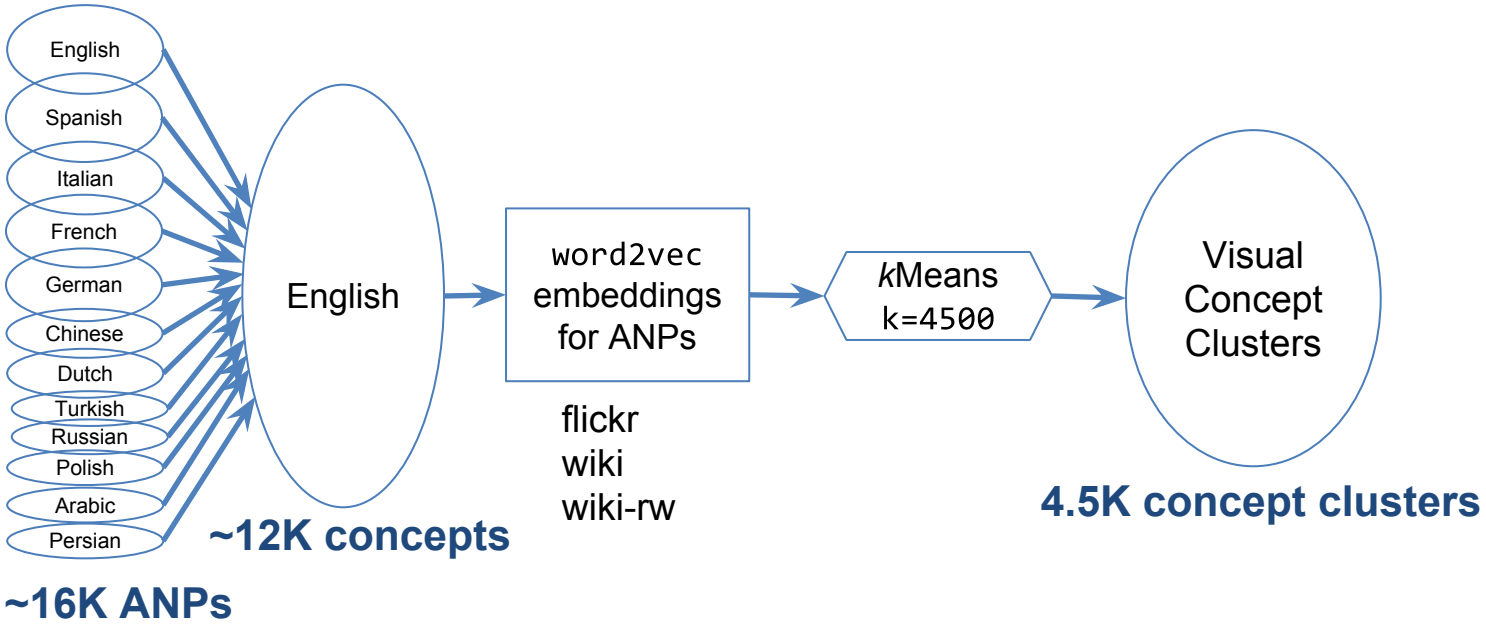


ENGLISH: **healthy coffee**

CONCEPT CLUSTERING

Approximate Multilingual Concept Matching

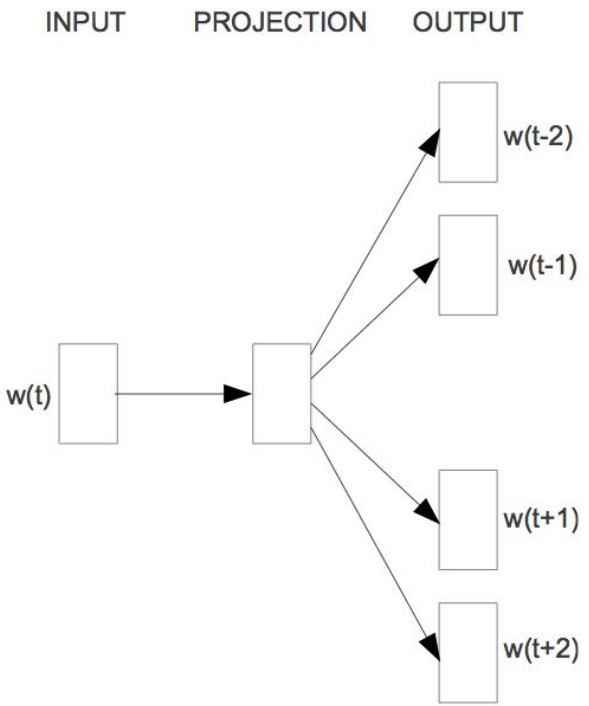
Single-stage: Use embeddings that are directly learned keeping ANPs as single tokens



k value is decided using inertia, sentiment and semantic consistency

Word Embedding Model

- Skip-gram model (`word2vec`)[1]
  - Google News 100B
  - Wikipedia 1.74B
  - Wikipedia + Reuters + WSJ 1.96B
  - Flickr 100 Million 0.75B
- Concept vectors
  - Sum of words composition
  - Directly learned (ANPs as tokens)



[1] Tomas Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado and Jeffrey Dean
Distributed Representations of Words and Phrases and their Compositionality
NIPS, Lake Tahoe, Nevada, USA, 2013

Approximate Concept Matching: Two-stage

- **Noun-first clustering:** concepts that talk about similar objects
- **Adjective-first clustering:** concepts about closely related emotions
- Ontologies to easily explore the dataset

Noun-first clustering

yard flowers summer spring lawn garden

romantic garden ecological garden **beautiful garden** celestial garden

rainy spring rainy summer

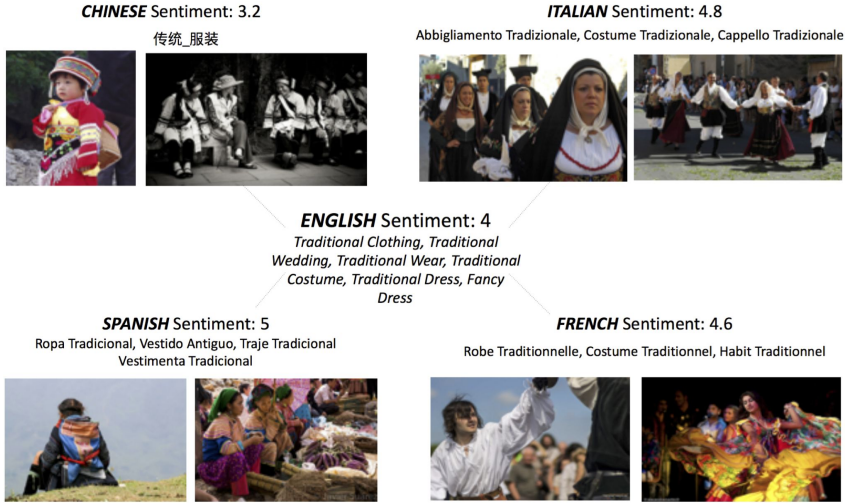
Adjective-first clustering

joyous beautiful festive floral happy delightful

delightful roses beautiful flowers **beautiful garden** beautiful butterfly

happy wedding happy marriage

We matched multilingual concepts…



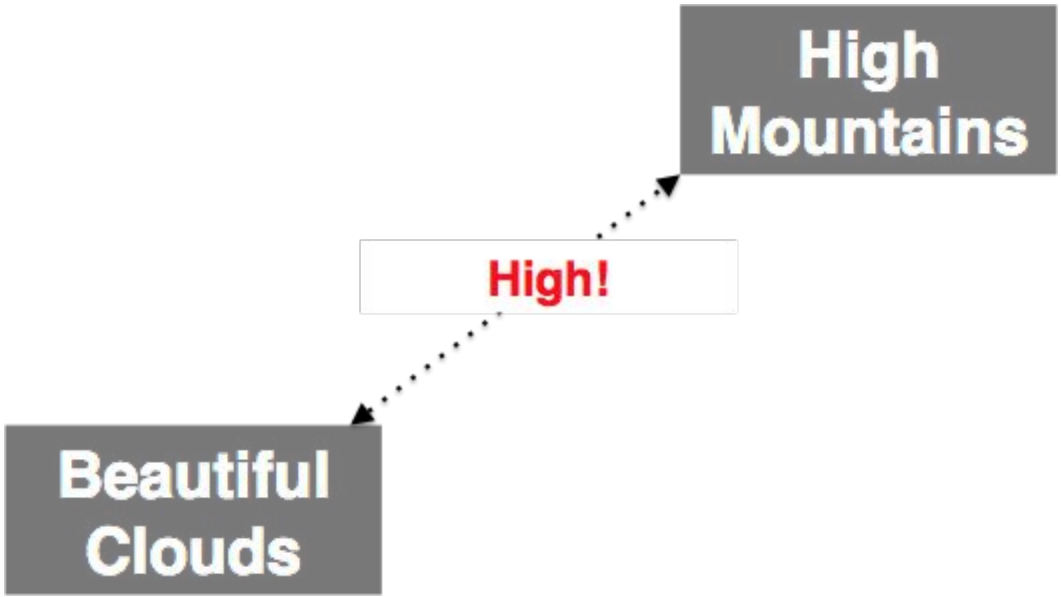
… but how do we evaluate the clustering methods?

- Semantic consistency
- Sentiment consistency

EVALUATION

SEMANTIC CONSISTENCY

Clustering Evaluation: Visual semantic relatedness



Semantic distance

Clustering Evaluation: Visual semantic relatedness

Visually-grounded semantic distance

Clustering Evaluation: Visual semantic relatedness

- How often do two visual concepts appear together?
  - **Tag co-occurrence matrix (n × n)**
- ANPs can be described as
  - **Co-occurrence vectors h_i, h_j in R^n**
    - n is the number of translated ANPs

- **Visual semantic distance between ANPs**

$$d(ANP_i, ANP_j) = 1 - cosine(h_i, h_j)$$

Clustering Evaluation: Semantic consistency

Visual **Semantic** Relatedness for different clustering methods

For each clustering method:

$$\text{sem}_C = \frac{1}{C}\sum_{c=1}^{C}\frac{\sum_{j:j\neq i\ \&\ Uij\neq 0}^{|\{i,\ldots,N_c\}|} d(\text{ANP}_{c,i}, \text{ANP}_{c,j})}{N_c}$$

Average over all clusters

Average visual semantic distance in a cluster for all ANP pairs whose semantic distance is greater than 0

C = number of non-unary clusters
Nc = number of ANPs for a cluster c

Inter-cluster distance was not significantly different

EVALUATION

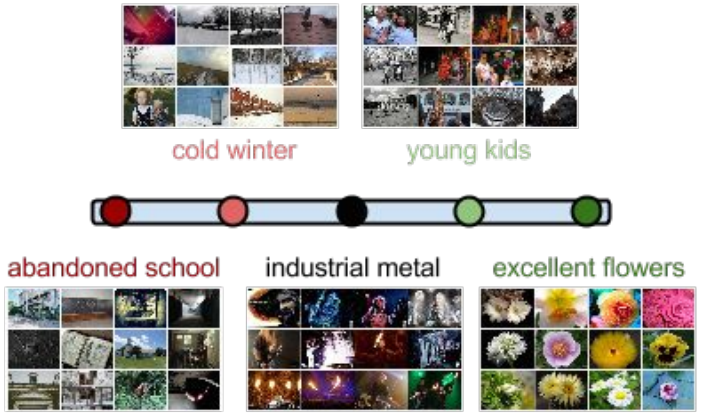
SENTIMENT CONSISTENCY

Clustering Evaluation: Visual sentiment of concepts

Visual **Sentiment** Consistency for different clustering methods

MULTIMODAL CROWDSOURCING EXPERIMENT

- 11 languages
- Native speakers
- Five grades
- Multimodal: Text + Images



Clustering Evaluation: Sentiment consistency

Visual **Sentiment** Consistency for different clustering methods

For each clustering method:

$$\text{sen}_C = \frac{1}{C}\sum_{c=1}^{C}\frac{\sum_{i=1}^{N_c}(\text{sen}(\text{ANP}_{c,i}) - \text{sen}_c)^2}{N_c}$$

Average sentiment in a cluster

Average over all clusters

Average visual sentiment error in a cluster

C = number of non-unary clusters
Nc = number of ANPs for a cluster c

EVALUATION

RESULTS

Clustering Evaluation: Results on Full Corpus

Single-step clustering performs better than two-step clustering

Directly learned ANP representations better than word-based ones

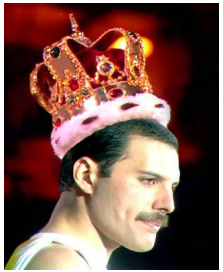
Method	Embeddings	Sentiment Cons.	Semantic Cons.	Overall Cons.
2-stage_noun	gnews (w=5)	0.278	0.676	0.477
2-stage_adj	gnews (w=5)	**0.161**	0.614	0.388
1-stage	wiki-anp (w=10)	0.239	0.659	0.449
1-stage	wiki_rw-anp (w=10)	0.242	0.582	0.412
1-stage	flickr-anp (w=10)	0.242	**0.535**	**0.388**
1-stage	wiki-anp (w=5)	0.239	0.659	0.449
1-stage	wiki_rw-anp (w=5)	**0.234**	0.579	0.407
1-stage	flickr-anp (w=5)	0.246	0.532	0.389

Application: Portrait concept clustering

Pictures of people are different from other photographs.

- Faces grasp human attention more than other subjects (neuroscience, computational social science)
- Eastern and Western Languages assign emotions differently (psychology theory)

Gorgeous girl



Grandi Persone

Ojos Lindos

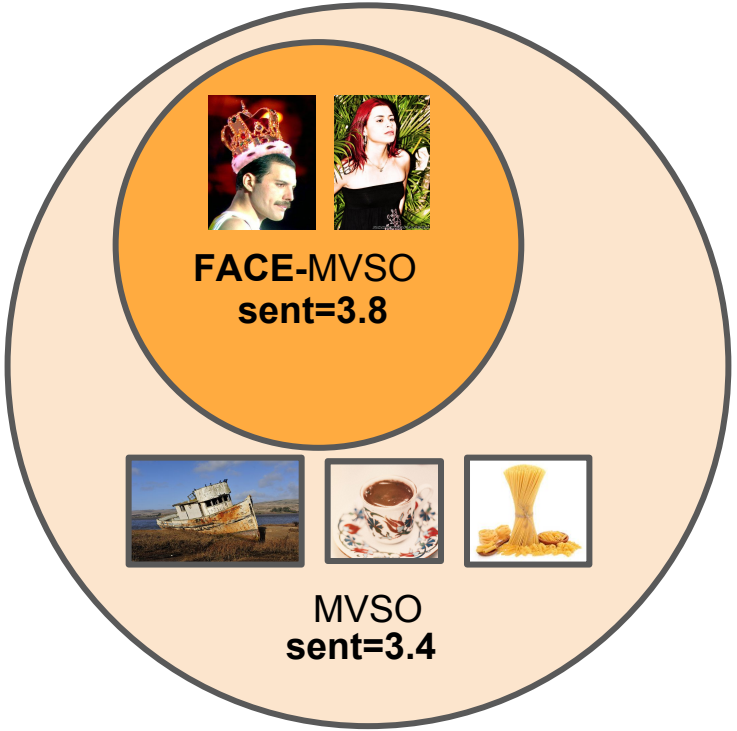
Regarde Triste

Güzel Kız

Application: Portrait concept clustering

Portrait-Based Sentiment Ontology using Face Detection

- Face ANPs (~2K, 3M images) have higher sentiment!
- **Highest** sentiment difference: **Chinese** 3.6 → 4.3 (+~20%)
- **Lowest** sentiment difference: **Turkish** 3.6 → 3.5 (-0.3%)



Clustering Evaluation on Face-ANPs: Results

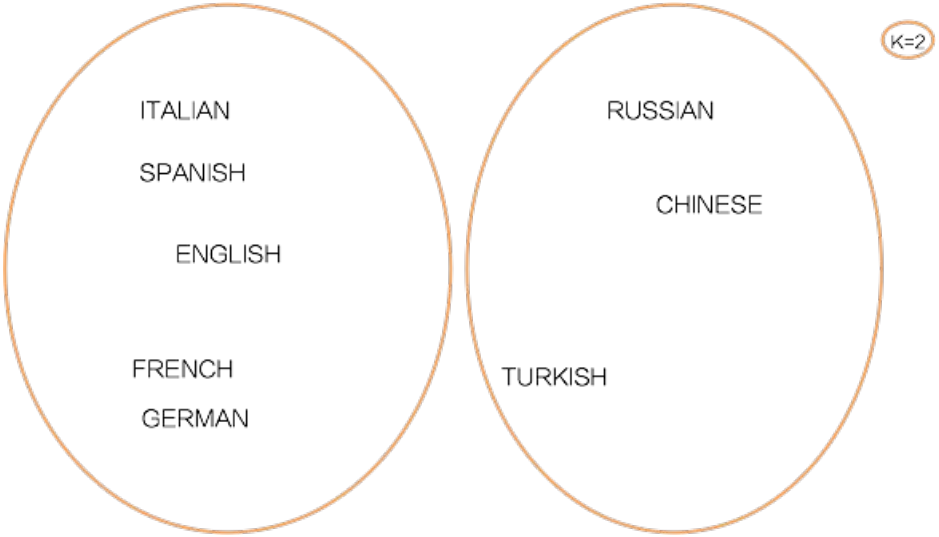
- Similar results as full corpus
- **Clusters with more languages → Higher sentiment!**
- Different Sentiment for different languages (Eastern vs. Western)

Method	Embeddings	Sentiment Cons.	Semantic Cons.	Overall Cons.
2-stage_noun	wiki (w=5)	0.534	0.586	0.56
2-stage_noun	wiki_rw (w=5)	0.510	0.614	0.562
2-stage_noun	flickr (w=5)	0.526	**0.513**	0.519
2-stage_noun	gnews (w=5)	0.309	0.569	0.439
2-stage_adj	wiki (w=5)	0.581	0.930	0.755
2-stage_adj	wiki_rw (w=5)	0.472	0.560	0.516
2-stage_adj	flickr (w=5)	0.455	0.519	0.487
2-stage_adj	gnews (w=5)	**0.178**	0.522	**0.350**
1-stage	wiki-anp (w=10)	**0.240**	0.576	0.408
1-stage	wiki_rw-anp (w=10)	0.257	0.508	0.382
1-stage	flickr-anp (w=10)	0.262	**0.489**	**0.375**
1-stage	wiki-anp (w=5)	0.250	0.583	0.416
1-stage	wiki_rw-anp (w=5)	0.281	0.522	0.402
1-stage	flickr-anp (w=5)	0.280	0.502	0.391

Application: Portrait concept clustering

Which languages are most similar when talking about faces?

Language representation: distribution of ANPs over 1000 clusters



Two clusters:

Eastern vs. Western

As seen in previous psychology studies

Application: Portrait concept clustering

Which languages are most similar when talking about faces?

Two clusters:

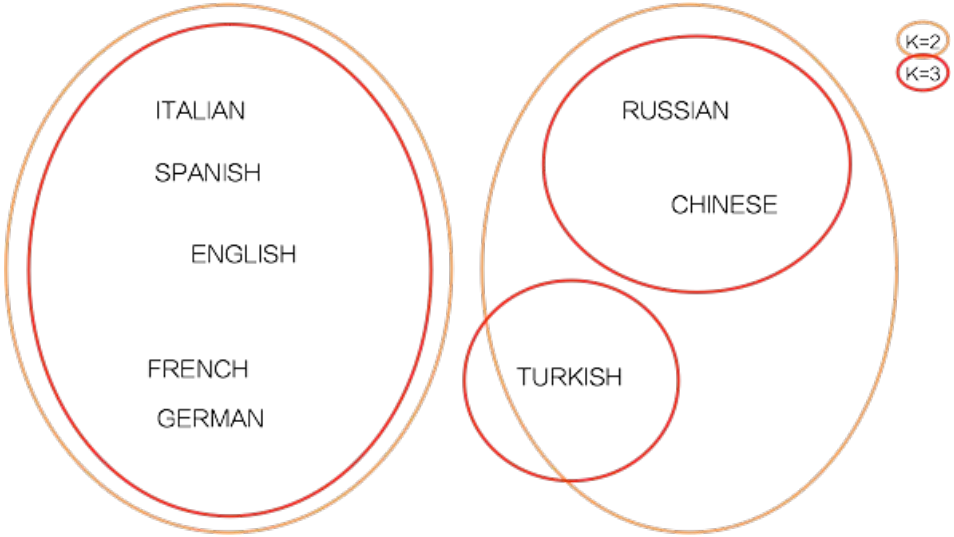
Eastern vs. Western

As seen in previous psychology studies

Application: Portrait concept clustering

Which languages are most similar when talking about faces?

Language representation: distribution of ANPs over 1000 clusters



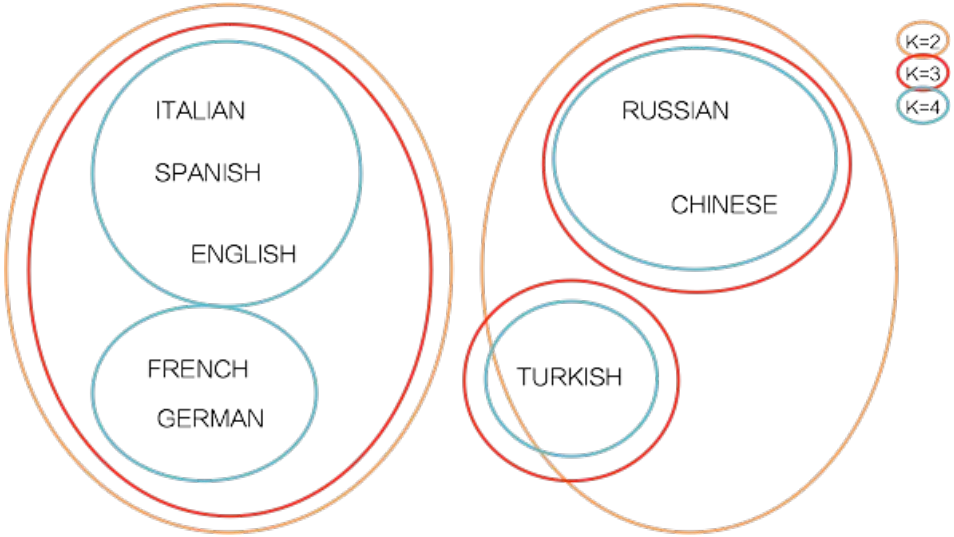
Three clusters:

Turkish detaches from the Eastern cluster

Application: Portrait concept clustering

Which languages are most similar when talking about faces?

Language representation: distribution of ANPs over 1000 clusters



Four clusters:

French/German

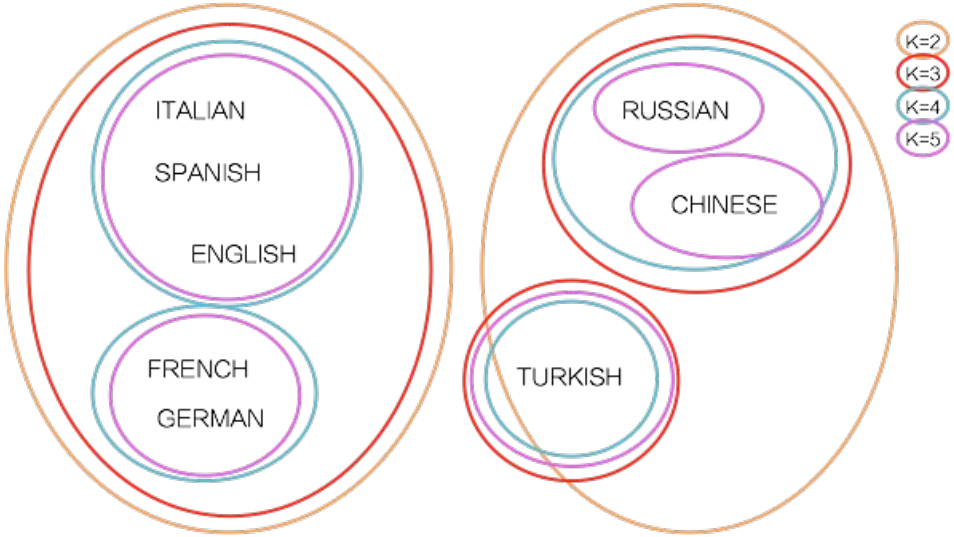
VS

Italian/Spanish/English

Application: Portrait concept clustering

Which languages are most similar when talking about faces?

Language representation: distribution of ANPs over 1000 clusters



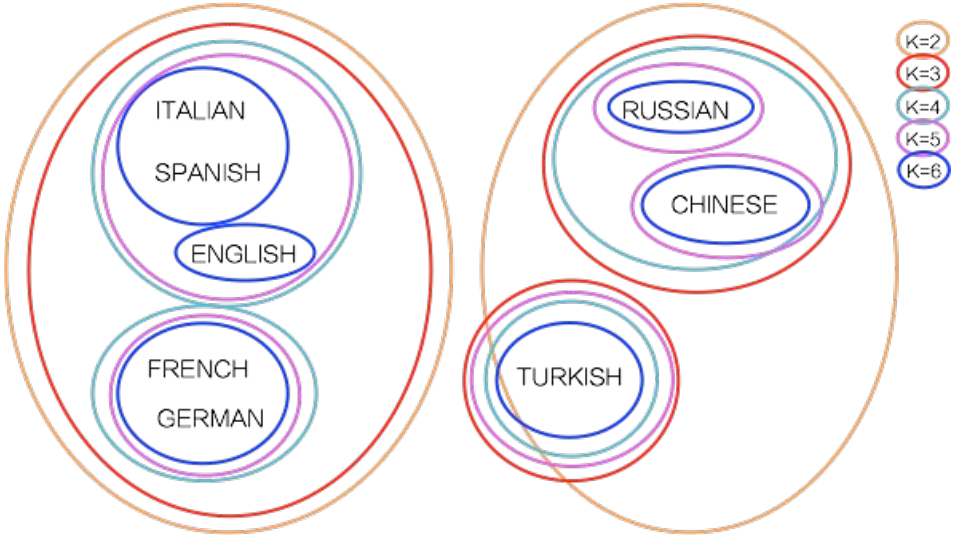
Five clusters:

Three Eastern languages are separated

Application: Portrait concept clustering

Which languages are most similar when talking about faces?

Language representation: distribution of ANPs over 1000 clusters



Six clusters:

Italian stays with Spanish

French with German

English as a single cluster

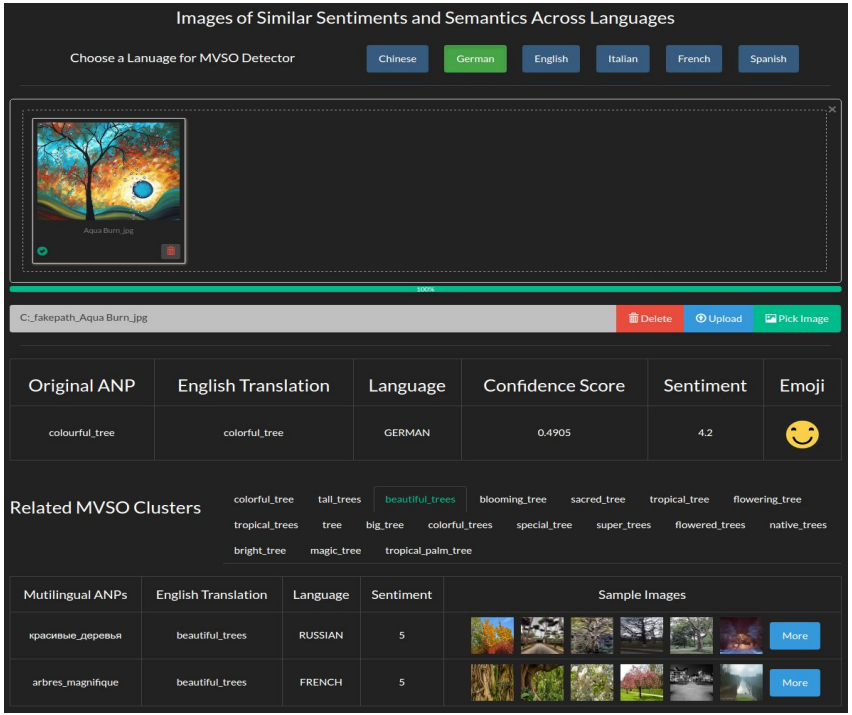
Summary

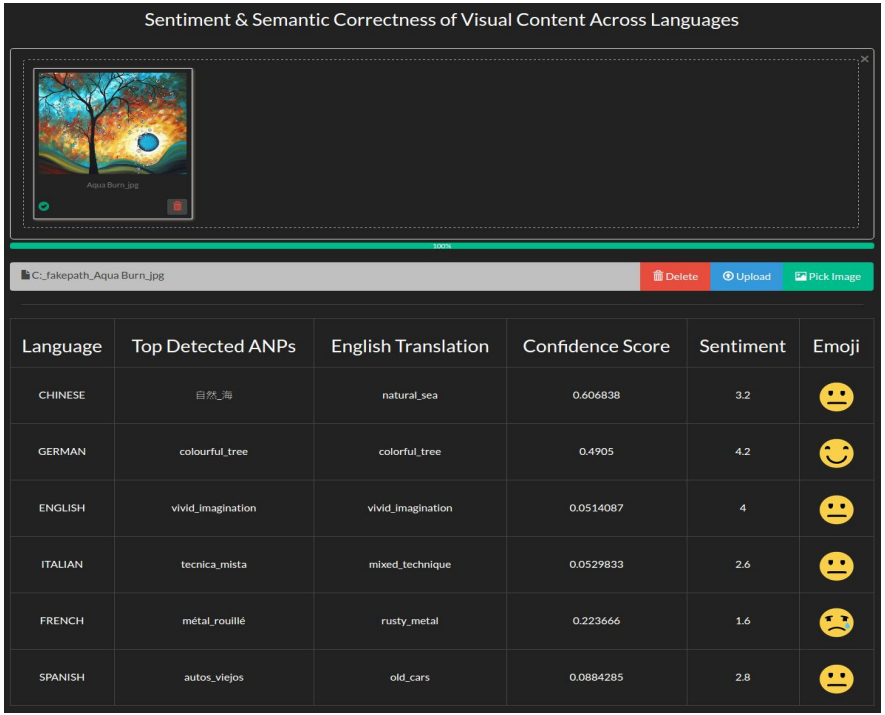
- **Domain consistency**
  - Word embeddings trained on a visually grounded corpus (Flickr) improve cluster quality for ANPs mined from visually grounded data
- **Single-token clustering**
  - Clustering adjectives noun pairs as single tokens proved merit
- **Visual semantic relatedness**
  - Measuring relatedness by tag co-occurrence is an effective evaluation for semantic visual grounding
- **Crowdsourced ANP sentiment**
  - Gathered a crowdsourced dataset of multimodal sentiment by ANPs
- **Eastern vs. Western**
  - We automatically discovered interesting and intuitive cultural differences

Demo

Complura: Exploring and Leveraging a Large-scale Multilingual Visual Sentiment Ontology

http://mvso.cs.columbia.edu/complura.html



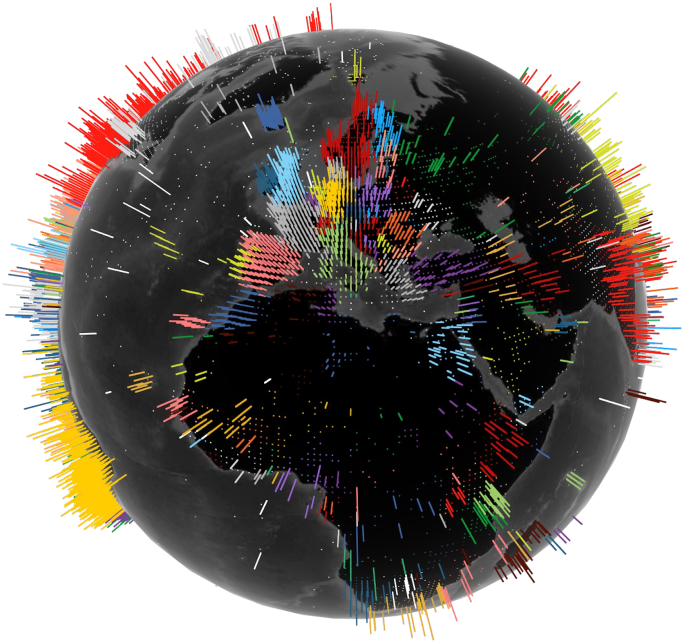


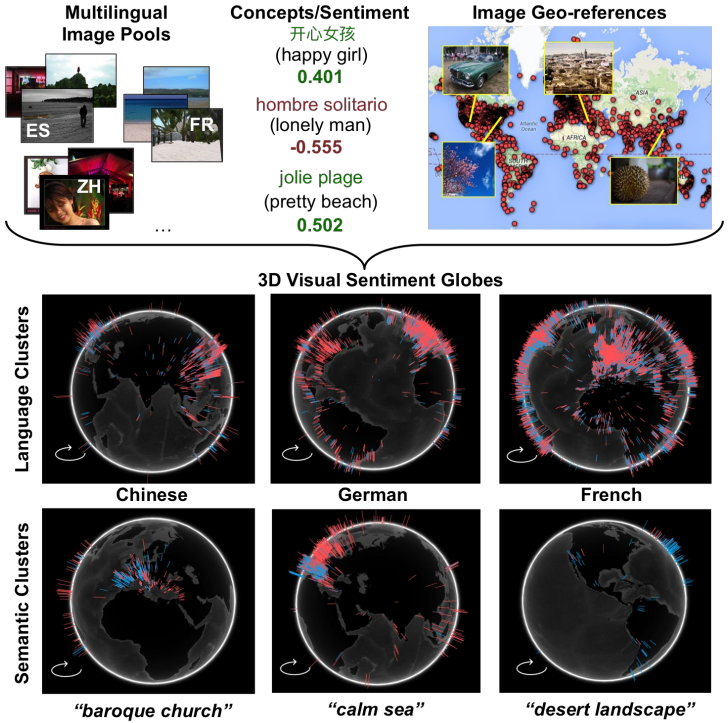
Visit the demo sessions for a live demo!

Demo

SentiCart: Cartography and Geo-contextualization for Multilingual Visual Sentiment

http://mvso.cs.columbia.edu/senticart.html





Visit the demo sessions for a live demo!

Thank you for your interest and questions!

For contacts and download links:
http://mvso.cs.columbia.edu

Question: What's Next?

- Use semantically aligned representations instead of translating to pivot
- Visually align ANP representations based on tag co-occurrence
- Improve detection, visual sentiment prediction and recommendation